



JP2001067187

Biblio Page 1

esp@cenet



## STORAGE SUB-SYSTEM AND ITS CONTROL METHOD

Patent Number: JP2001067187  
Publication date: 2001-03-16  
Inventor(s): ARAKAWA TAKASHI; MOGI KAZUHIKO; YAMAKAMI KENJI; ARAI HIROHARU  
Applicant(s):: HITACHI LTD  
Requested Patent: ☐ JP2001067187 (JP01067187)  
Application Number: JP19990242713 19990830  
Priority Number(s):  
IPC Classification: G06F3/06 ; G06F12/00  
EC Classification:  
Equivalents:

### Abstract

**PROBLEM TO BE SOLVED:** To simplify a work for optimizing arrangement by re-arrangement by the user of a disk array system or the like by changing the correspondence of a logical storage area from a physical storage area into the second physical storage area and executing re-arrangement.

**SOLUTION:** A control part 300 automatically executes re-arrangement execution processing at the set time and date. That is, the part 300 copies contents stored in a re-arrangement source physical area in a re-arrangement destination physical area based on re-arrangement information 408. Moreover, at the point of time when the copying is completed and the whole contents of the re-arrangement source physical area are reflected in the re-arrangement destination physical area, the control part 300 changes a physical area corresponding to a logical area for executing re-arrangement in logical/physical correspondence information 400 from the re-arrangement source physical area into the re-arrangement destination physical area. Besides, the control part 300 uses the re-arrangement destination physical area on a non-usage physical area 1470, changes the re-arrangement source physical area into the non-usage one and, moreover, updates the time and date of re-arrangement execution time information 406 into the one for a next time by referring to time and date updating information on re-arrangement execution time information 406.

Data supplied from the esp@cenet database - I2

## (19) 日本国特許庁 (J P)

## (12) 公開特許公報 (A)

(11) 特許出願公開番号  
特開2001-67187  
(P2001-67187A)

(43) 公開日 平成13年3月18日(2001.3.18)

| (51) Int.Cl. |       | P I    |       | 分類記号    |         |
|--------------|-------|--------|-------|---------|---------|
| G 06 F       | 3/06  | G 06 F | 3/06  | 3 01 A  | 5 B 065 |
|              |       |        |       | 5 4 0   | 5 B 082 |
|              | 12/00 |        | 12/00 | 5 0 1 B |         |

審査請求 未請求 請求項の数10 O L (全 24 項)

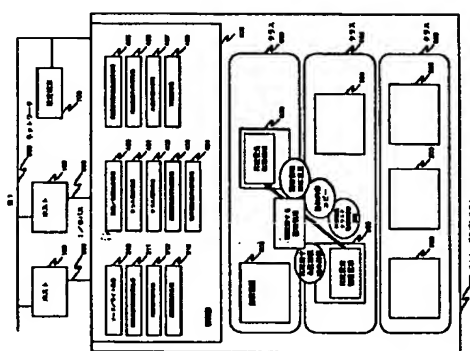
|           |                       |          |  |
|-----------|-----------------------|----------|--|
| (21) 出願番号 | 特願平11-242713          | (71) 出願人 | 00005108<br>株式会社日立製作所  |
| (22) 出願日  | 平成11年8月30日(1999.8.30) | (72) 発明者 | 株式会社日立製作所 神田駿河台四丁目6番地<br>荒川 敬史<br>株式会社日立製作所 神田駿河台四丁目6番地 株<br>神奈川県川崎市麻生区玉環寺1099番地 株<br>株式会社日立製作所 システム開発研究所内<br>渡本 和彦<br>株式会社日立製作所 麻生区玉環寺1099番地 株<br>株式会社日立製作所 システム開発研究所内<br>100075056<br>(74) 代理人 弁理士 作田 廣夫 |

## (54) 【発明の名称】 ストレージサブシステム及びその制御方法

## (57) 【要約】

【要約】ストレージサブシステムのユーザまたは保守員が記憶領域の物理的再配置による配置最適化を行うための作業を簡便にするストレージサブシステムおよび制御方法を提案する。

【解決手段】ストレージサブシステム200は、記憶装置500を、それぞれ属性を有する複数の組(クラス)600として管理し、クラス属性に基づき好適な再配置先のクラスを決定する。



## 【特許請求の範囲】

【請求項1】 複数の記憶装置と、前記記憶装置の使用状況情報を取得する手段と、前記計算機がリードライト対象とする論理記憶領域と前記記憶装置の第一の物理記憶領域とを対応づけを行う手段とを有し、1台以上の計算機に接続するストレージサブシステム(クラス)に分類され、前記記憶装置は複数の組(クラス)に分かれ、前記記憶装置は設定された属性を有し、前記ストレージサブシステムは、前記使用状況情報および前記クラス属性に基づき、前記記憶装置から再配置する論理記憶領域を決定し、前記記憶装置の再配置先として利用可能な第二の物理記憶領域を前記クラス内から選択し、前記第一の物理記憶領域の内容を前記第二の物理記憶領域にコピーするとともに論理記憶領域の対応づけを前記第一の物理記憶領域から前記第二の物理記憶領域へ変更して再配置を行うことを特徴とするストレージサブシステムの制御方法。

【請求項2】 請求項1に記載のストレージサブシステムは、前記の制御方法であって、ストレージサブシステムは、前記使用状況情報を蓄積し、設定された期間の前記使用状況情報に基づき、論理記憶領域の再配置先を決定し、設定された時間に再配置を行うことを特徴とするストレージサブシステムの制御方法。

【請求項3】 請求項1または2に記載のストレージサブシステムの制御方法であって、ストレージサブシステムは、使用状況情報として、記憶装置の単位時間当たりの使用時間(使用率)を用い、各クラスは、属性として設定されたクラス間の性能順位と使用率上限値を有し、前記ストレージサブシステムは、クラスの使用率上限値を超えている記憶装置から再配置する論理記憶領域を選択し、前記記憶装置の再配置先のクラスを前記順位の上位のクラスから、各クラスの使用率上限値を超えないように決定することを特徴とするストレージサブシステムの制御方法。

【請求項4】 請求項1または2に記載のストレージサブシステムの制御方法であって、ストレージサブシステムは、使用状況情報として、記憶装置の単位時間当たりの使用時間(使用率)を用い、各クラスは、属性として設定されたクラス間の性能順位と使用率上限値を有し、前記ストレージサブシステムは、クラスの使用率上限値を超えている記憶装置から再配置する論理記憶領域を選択し、前記記憶装置の再配置先として利用可能な物理記憶領域を同一クラス内の記憶装置から、前記クラスの使用率上限値を超えないように決定することを特徴とするストレージサブシステムの制御方法。

【請求項5】 請求項1または2に記載のストレージサブシステムの制御方法であって、ストレージサブシステムは、使用状況情報として、記憶装置の単位時間当たりの使用時間(使用率)を用い、各クラスは属性として設定された対象アクセス種別と使用率上限値を有し、前記ス

トレージサブシステムは、クラスの使用率上限値を超えている記憶装置から再配置する論理記憶領域を選択し、前記記憶装置の再配置先のクラスを前記分析結果に基づいて前記記憶装置の再配置先のクラスを前記対象アクセス種別のクラスとして決定することを特徴とするストレージサブシステムの制御方法。

【請求項6】 1台以上の計算機に接続し、複数の記憶装置と、前記記憶装置の使用状況情報を取得する手段と、前記計算機がリードライト対象とする論理記憶領域と前記記憶装置の第一の物理記憶領域とを対応づけを行う手段とを有するストレージサブシステムであって、前記複数のディスク装置をそれぞれ属性を有する複数の組(クラス)として管理する手段と、前記使用状況情報および前記クラス属性に基づき、前記記憶装置から再配置する論理記憶領域を前記第一の物理記憶領域から前記第二の物理記憶領域へ変更して再配置を行う手段とを有することを特徴とするストレージサブシステム。

【請求項7】 請求項6に記載のストレージサブシステムであって、ストレージサブシステムは、前記使用状況情報を蓄積し、設定された期間の前記使用状況情報に基づき、論理記憶領域の再配置先を自動的に決定する手段と、設定された時間に再配置を行う手段とを有することを特徴とするストレージサブシステム。

【請求項8】 請求項6または7に記載のストレージサブシステムであって、ストレージサブシステムは、使用状況情報として記憶装置の単位時間当たりの使用時間(使用率)を用い、各クラスは、属性として設定されている使用率上限値を有し、前記ストレージサブシステムは、各クラスに属性として設定されている使用率上限値を超えている記憶装置から再配置する論理記憶領域を選択する手段と、前記記憶装置の再配置先のクラスを各クラスに属性として設定されているクラス間の性能順位から、各クラスの使用率上限値を超えないように決定する手段とを有することを特徴とするストレージサブシステム。

【請求項9】 請求項6または7に記載のストレージサブシステムであって、ストレージサブシステムは、使用状況情報として、記憶装置の単位時間当たりの使用時間(使用率)を用い、各クラスは属性として設定された対象アクセス種別と使用率上限値を有し、前記ストレージサブシステムは、属性として設定されたクラスの使用率上限値を超えている記憶装置から再配置する論理記憶領域を選択する手段と、前記記憶装置の再配置先のクラスを前記分析結果に基づいて各クラスの再配置先のクラスを前記分析結果に基づいて各クラスの再配置先のクラスを前記分析結果に基づいて各クラスの再配置先のクラスを超えないように決定する手段とを有することを特徴とする

の最速配置を実現するディスクアレイシステムが提示されている。

【0005】

【発明が解決しようとする課題】特開平9-27454号公報に示されるような従来の技術における配置最適化の実行方法については以下の課題がある。

【0006】再配置する論理記憶領域の選択および再配置先の物理記憶領域の選択にあたり、ディスクアレイシステムのユーザまたは保守員が、前記ディスクアレイシステムの構成や個々のディスク装置の特性や性能などの情報を確認して前記選択を行わなければならないが、ユーザまたは保守員による作業が煩雑となっていた。

【0007】また、ディスクアレイシステムが選択を自動的に行う場合においても、ユーザまたは保守員が前記個々のディスク装置の情報を確認して選択基準値を規定しなければならないが、やはりユーザまたは保守員による作業が煩雑となっていた。特に、上記のように異種のレベラや異種のディスク装置の存在するディスクアレイシステムについては情報管理の煩雑さが増大する。

【0008】また、ディスクアレイシステムが選択のために行う1/Oアクセス情報の参照は、ホストコンピュータおよびディスクアレイシステムを含むシステムで行われる処理のスケジューリングの特性を考慮してなかつた。一般にコンピュータシステムで行われる処理と処理とに伴う1/Oは、ユーザによって作成されたスケジューリングに則って行われており、また処理および1/Oの傾向は日毎、月毎、年毎などの周期性を示す場合も多く、一般にユーザは特定期間の処理および1/Oに関心があると考えられる。

【0009】また上記従来技術において、再配置による性能チューニング方法については以下の課題がある。物理的再配置による性能チューニング方法は、ディスク装置、すなわち、物理記憶領域の使用状況に変更を加えるものであるが、従来の技術においては、ホストコンピュータからの論理記憶領域に対する1/Oアクセスについての情報を参照するため、再配置する論理記憶領域の選択が行えない可能性があった。

【0010】また、ホストコンピュータからのシーケンシャルアクセスとランダムアクセスが顕著に、同一のディスク装置に含まれる別々の物理記憶領域に対して行われる場合でも、シーケンシャルアクセスとランダムアクセスを異ならせるディスク装置に分離するために、再配置先のディスク装置を任意に特定して自動的再配置を行わなければならない。一般に、ホストコンピュータからの処理要件として、データ長の小さいランダムアクセスには短時間で応答（高応答性能）が求められるが、同一ディスク装置にデータ長の大きいシーケンシャルアクセスが存在する場合、ランダムアクセスの応答時間はシーケンシャルアクセスの処理に阻害されて長くなり、

ランダムアクセスである。

【請求項10】請求項6、7、8、または9に記載のストレージサブシステムであって、ストレージサブシステムは、複数のディスク装置を行なうディスクアレイであり、前記ディスク装置の使用状況情報として用いる手段を有することを特徴とするストレージサブシステム。

【発明の詳細な説明】

【0001】  
【発明の属する技術分野】本発明は、複数の記憶装置を有するストレージサブシステム、およびその制御方法に関する。

【0002】

【従来の技術】コンピュータシステムにおいて、高性能を実現する二重記憶システムの1つにディスクアレイシステムがある。ディスクアレイシステムは、複数のディスク装置をアレイ状に配置し、前記各ディスク装置に分割格納されるデータのリード/ライトを、前記各ディスク装置を並列に動作させることによって、高速に行うシステムである。ディスクアレイシステムに関する論文としては、D. A. Patterson, G. Gibbons, and K. H. Kats, "A Case for Redundant Arrays of Inexpensive Disks (RAID)" (in Proc. ACM SIGMOD, pp. 109-116, June 1988)がある。この論文では、冗長性を付加したディスクアレイシステムに対し、その構成に応じてレベル1からレベル5の種別を定義している。これらの種別に加えて、冗長性無し（ディスクアレイシステムをレベル0と呼ぶこともある。上記の各レベルは冗長性などにより実現するためのコストや性能特性などが異なるため、ディスクアレイシステムを構築するにあたって、複数のレベルのアレイ（ディスク装置の組）を混在させることも多い。ここでは、この組のことをパーティティクルと呼ぶ。

【0003】ディスク装置は、性能や容量などによりコストが異なり、ディスクアレイシステムを構築するために必要に応じて異なる複数のディスク装置を用い、または性能や容量の異なる複数のディスク装置を用いることがある。

【0004】ディスクアレイシステムに格納されるデータを上記のようにディスク装置に分散して配置したため、ディスクアレイシステムに格納されるホストコンピュータがアクセスする論理記憶領域とディスク装置の記憶領域を示す物理記憶領域の対応づけ（アドレス変換）を行う。特開平9-27454号公報には、ホストコンピュータからの論理記憶領域に対する1/Oアクセスについての情報を取得する手段と、論理記憶領域の物理記憶領域への対応づけを変更して物理的再配置を行う手段により、格納されたデータ

応答性能は悪化してしまう。

【0011】本発明の第一の目的は、ディスクアレイシステムのユーザまたは保守員が再配置による配置最適化を行うための作業を簡便にすることにある。

【0012】本発明の第二の目的は、ホストコンピュータおよびディスクアレイシステムを含むシステムでの処理のスケジューリングを考慮した再配置による配置最適化を可能にすることにある。

【0013】本発明の第三の目的は、再配置する論理記憶領域の選択および再配置先の物理記憶領域の選択にあたり、実際の記憶装置であるディスク装置の使用状況に基づき選択を行う、ディスクアレイシステムの制御方法およびディスクアレイシステムを提供することにある。

【0014】本発明の第四の目的は、ディスクアレイシステムにおける同一ディスク装置での異なるシーケンシャルアクセスとランダムアクセスの存在に対し、再配置先のディスク装置を任意に特定して再配置によりシーケンシャルアクセスおよびランダムアクセスを異なるディスク装置に自動的に分離することができようようにすることにある。

【0015】  
【課題を解決するための手段】上記の第一の目的を実現するために、1台以上のホストコンピュータに接続するディスクアレイシステムは、配下の複数のディスク装置の使用状況情報を取得する手段と、ホストコンピュータがリード/ライト対象とする論理記憶領域とディスク装置の第一の物理記憶領域との対応づけを行う手段とを有し、さらに、複数のディスク装置をそれぞれ属性を有する複数の組（クラス）として管理する手段と、使用状況情報およびクラス属性に基づき論理記憶領域に好適な再配置先のクラスを決定する手段と、論理記憶領域の再配置先として利用可能な第一の物理記憶領域をクラス内から選択する手段と、第二の物理記憶領域の内容を前記第一の物理記憶領域にコピーするとともに論理記憶領域の対応づけを第一の物理記憶領域から第二の物理記憶領域へ変更して再配置を行う手段を備える。

【0016】また、上記第二の目的を実現するために、ディスクアレイシステムは、使用状況情報を蓄積し、設定された期間の使用状況情報に基づき、論理記憶領域の再配置先を決定する手段と、設定された時間に再配置を行う手段を備えることができる。

【0017】また、上記第三の目的を実現するために、ディスクアレイシステムは、使用状況情報として、ディスク装置の単位時間当たりの使用時間（使用率）を用いる手段を備える。

【0018】また、上記第四の目的を実現するために、ディスクアレイシステムは、各クラスに属性として設定された対象アクセス種別（シーケンシャル/ランダムアクセス種別）と使用率上限値を用いて、クラスの使用率上限値を超えている記憶装置から再配置する論理記憶領域

域を選択し、論理記憶領域に対するアクセス種別の分析結果に基づいて論理記憶領域の再配置先のクラスを好適なアクセス種別のクラスから、各クラスの使用率上限値を超えないように決定する手段を備える。

【0019】

【発明の実施の形態】以下、本発明の実施の形態を図1〜図27を用いて説明する。

【0020】<第一の実施の形態>本実施の形態では、クラス600に基づく再配置の判断と、再配置判断および実行のスケジューリングについて説明する。

【0021】図1は、本発明の第一の実施の形態における計算機システムの構成図である。

【0022】本実施の形態における計算機システムは、ホスト100、ストレージサブシステム200、制御部150を有している。

【0023】ホスト100は、ストレージサブシステム200に1/Oバス800を介して接続し、ストレージサブシステム200に対しリード/ライトの1/Oを行う。1/Oの際、ホスト100は、ストレージサブシステム200の記憶領域について論理記憶領域を指定する。1/Oバス800の例としては、ECON、SCSI、ファイバチャネルなどがある。

【0024】ストレージサブシステム200は、制御部300および複数の記憶装置500を有する。制御部300は、リード/ライト処理310、使用状況情報取得処理311、再配置判断処理312、及び再配置実行処理313を行う。また、ストレージサブシステム200は、論理/物理対応情報400、クラス構成情報401、クラス属性情報402、論理記憶領域使用状況情報403、物理記憶領域使用状況情報404、再配置判断対象期間情報405、再配置実行時間情報406、未使用記憶領域情報407、及び再配置情報408を保持する。

【0025】ホスト100、制御部300、および制御部700は、ネットワーク900で接続される。ネットワーク900の例としては、FDDI、ファイバチャネルなどがある。

【0026】ホスト100、制御部300、および制御部700には、各々での処理を行うためのメモリ、CPUなど、計算機において一般に用いられる構成要素もそれぞれ存在するが、本実施の形態の説明においては重要でないため、ここでは説明を省略する。

【0027】ホスト100が、ストレージサブシステム200に対してリード/ライトを行う場合のリード/ライト処理310、および使用状況情報取得処理311について図2で説明する。

【0028】リード/ライト処理310において、ホスト100は、ストレージサブシステム200の制御部300に対しリードまたはライトを論理記憶領域を指定して要求する（ステップ1000）。要求を受領した制御部300は、論理/物理対応情報400を用いて論理記憶領域

対応する物理領域を求め、すなわち論理領域のアドレス（論理アドレス）を物理領域のアドレス（物理アドレス）に変換する（ステップ110）。続いて制御部300は、リードノードの場合は、この物理アドレスの記憶装置500からデータを読み出してホスト100に転送し、ライトの場合は、ホスト100から転送されたデータを前記物理アドレスの記憶装置500に格納し（ステップ120）、さらに後述の使用状況情報取得処理311を行う。リードノード要求およびデータ転送は1/0バス800を介して行われる。

【0029】論理/物理対応情報400の一例を図3に示す。論理アドレスはホスト100がリードノードとして物理310で用いる論理領域を示すアドレスである。物理アドレスは実際にデータが格納される記憶装置500上の領域を示すアドレスであり、記憶装置番号および記憶装置500を示す。記憶装置番号は個々の記憶装置500内の記憶領域を示すアドレスである。

【0030】次に、使用状況情報取得処理311において制御部300は、リードノード/ライト処理310においてリードノード/ライト対象となった論理領域についての論理領域使用状況情報403と、リードノード/ライト処理310で実際に物理領域についての物理領域使用状況情報404を更新する（ステップ103、104）。論理領域使用状況情報403および物理領域使用状況情報404は、例えば使用頻度、使用率、リードノード/ライトに関する属性など、各々の論理領域と物理領域の各日時の使用状況に関する情報である。論理領域使用状況情報403および物理領域使用状況情報404の具体的な例は、以下の実施形態で説明する。

【0031】次に、制御部300が行う再配置判断処理312について図4で説明する。

【0032】記憶装置500は、ユーザによって、または初期状態として複数の組（クラス600）に分類されており、クラス600への分類はクラス構成情報401に設定されている。さらに、各クラス600は、ユーザによって、または初期条件として属性を設定されており、属性は、クラス属性情報402に設定されている。クラス属性情報402は、許容使用状況や好適な使用状況やクラス間優先順位などの属性に関する情報である。クラス構成情報401およびクラス属性情報402の具体的な例は、以下の実施形態で説明する。再配置判断対象期間情報405は、ユーザによってまたは初期条件として再配置判断処理312の対象とする使用状況情報の期間と期間更新情報406が設定されている。

【0033】再配置判断対象期間情報405の一例を図5に示す。開始1時から終了1時までの期間が対象期間となる。期間更新情報は次の対象期間の設定条件であり、例えば毎週、毎日、X時間後などがありうる。制御部300は、対象期間の論理領域使用状況情報403および物理領域使用状況情報404に基づき再配置

および物理領域使用状況情報404を参照し（ステップ110）、クラス属性情報402の各クラス600の許容使用状況などと比較して（ステップ111）、物理的再配置を行うべき論理領域を選択する（ステップ112）。

【0034】さらに、制御部300は、クラス属性情報402の許容使用状況や好適な使用状況やクラス間優先順位などを参照して（ステップ113）、論理領域の再配置先のクラス600を選択し（ステップ114）、さらに、クラス600に属する記憶装置500の中から論理領域の再配置先として未使用の物理領域を選択し（ステップ115）、選択結果を再配置情報408に出力する（ステップ116）。

【0035】再配置情報408の一例を図6に示す。論理領域は、再配置する論理領域であり、再配置元物理領域は、論理領域に対応する現在の物理領域を示す記憶装置番号と記憶装置内アドレスであり、再配置先物理領域は、再配置先の物理領域を示す記憶装置番号と記憶装置内アドレスである。図6に示すように再配置の立案は一つ以上行われる。さらに制御部300は、再配置判断対象期間情報405の期間更新情報を参照して、再配置判断対象期間情報405の対象期間を次回分を更新する（ステップ117）。上記処理において制御部300は、論理/物理対応情報400を用い、また前記の未使用の物理領域の検索に未使用領域情報407を用いる。

【0036】未使用領域情報407の一例を図7に示す。記憶装置番号は個々の記憶装置500を示す。記憶装置内アドレスは記憶装置500内の領域を示すアドレスである。記憶装置番号および装置内アドレスは物理領域を示し、使用/未使用の項目は、物理領域の使用/未使用の区別を示す。制御部300は、通常、再配置判断処理312を対象期間以後、後述の再配置実行処理313以前に自動的に実行する。

【0037】次に、制御部300が行う再配置実行処理313について図8で説明する。

【0038】再配置実行時刻情報406にはユーザによってまたは初期条件として再配置実行処理313を行う日時と日時更新情報が設定されている。

【0039】再配置実行時刻情報406の一例を図9に示す。制御部300は、設定された日時に以下に説明する再配置実行処理313を自動的に実行する。日時更新情報は次の再配置実行処理313を行う日時の設定条件であり、例えば毎週、毎日、X時間後などがありうる。制御部300は、再配置情報408に基づき再配置元物理領域に格納している内容を再配置先物理領域にコピーする（ステップ120）。さらに、コピーが完了して再配置元物理領域の内容が全て再配置先物理領域に反映された時点で、制御部300は、論理/物理対応情報400上の再配置を行う論理領域に対応する物理領域

を再配置元物理領域から再配置先物理領域に変更する（ステップ121）。

【0040】さらに、制御部300は、未使用物理領域470上の再配置先物理領域を使用し、再配置元物理領域を未使用に変更する（ステップ122）。さらに、制御部300は、再配置実行時刻情報406の日時更新情報を参照して、再配置実行時刻情報406の日時を次回分を更新する（ステップ123）。

【0041】ユーザまたは保守員は、制御部300が上記の処理で用いている各情報を、制御部300からネットワーク900を介して、またはホスト100からネットワーク900または1/0バス800を介して設定および確認すること、特に、再配置情報408を確認および設定して再配置案を修正や追加や削除などを行うことができる。

【0042】上記の処理を行うことによって、取得した使用状況情報および設定されたクラス属性に基づいて、ストレージサブシステム200において論理領域の物理的再配置を自動的に実行し、ストレージサブシステム200の最適化を行うことができる。さらに上記の再配置判断および実行の処理を繰り返して配置を修正していくことにより、使用状況の変動やその他の最適化の必要要因を吸収していくことができる。

【0043】特に、上記の処理により、ユーザまたは保守員は再配置による最適化を簡単に実行することができる。ユーザまたは保守員は、記憶装置500をクラス600の単位で管理できるため、記憶装置500の性能や信頼性や特性などの属性を個々の記憶装置500について管理する必要がある。さらに、ユーザまたは保守員は、記憶装置500の個々の属性が等しくない組に対しても、必要に応じて同一の属性を持つクラス600を設定して、1つの管理単位として扱うことができる。ただし、1つの記憶装置500が1つのクラス600を構成すると見なして1つの記憶装置500を管理単位として上記の再配置の処理を行うことも可能である。

【0044】また、ユーザまたは保守員は、ホスト100で行われる処理（ジョブ）の特性やスケジュールを考慮し、計算機システムで行われる処理と、この処理に伴う1/0は、ユーザによって作成されたスケジュールに則って行われる。ユーザは、特に最適化の対象としたい処理を有する場合、処理の期間を指定することが可能であり、本実施形態の期間を指定して再配置の処理によって、ストレージシステム200に行わせ、すなわち、前記期間の使用状況情報に基づいて上記の再配置による最適化を実現することができる。また、計算機システムで行われる処理および1/0の傾向は日毎、月毎、年毎などの周期性を示す場合も多い。特に、処理が定期処理に基づく処理である場合には、周期性が顕著となる。前述の場

合と同様にユーザは、図例において特に最適化対象として関心のある期間を指定して再配置による最適化を行うことができる。また、再配置実行処理313では、ストレージシステム200内で格納内容のコピーを行うが、ユーザはストレージシステム200があまり使用されていない時間帯にホスト100で実行されている処理の要求処理性能が低い期間を再配置実行処理313の再配置時刻として設定することで、ホスト100での要求処理性能が高い処理のストレージシステム200への1/0がコピーにより阻害されることを回避できる。

【0045】なお、記憶装置500は、それぞれ異なる性能、信頼性、特性や属性を持っているが、特に具体的には、磁気ディスク装置、磁気テープ装置、半導体メモリ（キャッシュ）のように異なる記憶媒体であってもよい。また、上記の例では未使用領域情報407は物理領域に基づいて記述されているとしたが、未使用の物理領域に対応する論理領域（論理アドレス）に基づいて記述されていてもよい。

【0046】<第二の実施形態>本実施形態では、使用状況情報としてのディスク装置使用率の適用と、クラス600の上限値およびクラス600間の性能順位による再配置判断について説明する。

【0047】図10は、本発明の第二の実施形態における計算機システムの構成図である。

【0048】本実施形態の計算機システムは、ホスト100、ディスクレイシステム201、制御部300を有している。本実施形態における計算機システムは、第一の実施形態でのストレージサブシステム200をディスクレイシステム201としたものに相当する。

【0049】ディスクレイシステム201は、制御部300とディスク装置502を有する。制御部300は、第一の実施形態での制御部300に相当する。ディスク装置502は、n台（nは2以上の整数）でRAID（ディスクアレイ）を構成しており、このn台のディスク装置502による組をパーティティグループ501と呼ぶ。RAIDの性質として、1つのパーティティグループ501に含まれるn台のディスク装置502は、n-1台のディスク装置502の格納内容から生成される冗長データが残り1台に格納されるといった冗長性上の関係を含めた格納内容が、並列動作性向上のためにn台のディスク装置502に分散格納されるなど、データ格納上の関係を持つ。この関係から各パーティティグループ501は動作上の1単位とみなすことができるが、冗長性や台数nなどにより実現するためのコストや特性などがあるため、ディスクレイシステム201を構成するにあたって、レールや台数nの異なるアレイ（パーティティグループ501）を混在させることも多く、またパーティティグループ501を構成するディスク装置502につ

いても、性能や容量などによりコストが異なるため、ディスクアレイシステム201を構成するにあたって最適なコストパフォーマンスを実現するために性能や容量の異なる複数のディスク装置502を用いることもあ  
55 る。よって本実施の形態においてディスクアレイシステム201を構成する各パーティグループ501は性能、信頼性、特性などの属性が同一であるとは限らず、特に性能について差異があるとする。

【0050】本実施の形態における論理/物理対応情報400の一例を図11に示す。

【0051】論理アドレスは、ホスト100がリード/ライト処理310で用いる論理領域を示すアドレスである。物理アドレスは実際にデータと前記冗長データが格  
10 納されるディスク装置502上の領域を示すアドレスであり、パーティグループ501と各々のディスク装置番号  
15 およびディスク装置内アドレスからなる。パーティグループ番号は個々のパーティグループ501を示す。ディ  
20 スク装置番号は個々のディスク装置502を示す。ディ  
25 スク装置内アドレスはディスク装置502内の領域を示すアドレスである。制御部300は、RAIDの動作  
として、冗長データに関する情報を前記リード/ライト  
30 処理310などで用いて処理するが、本実施の形態の説明では、パーティグループ502を動作上の1単位とし  
て説明するため、前記処理に關してはここでは特にふれ  
ない。

【0052】さらに第1の実施の形態と同様に、パーティグループ501は、ユーザによってまたは初期状態と  
35 して複数の組(クラス600)に分類されており、クラス600への分類はクラス構成情報401に設定されて  
いる。クラス構成情報401の一例を図12に示す。  
【0053】クラス番号は各クラス600を示す番号であ  
40 る。パーティグループ数は各パーティグループ番号は各  
45 クラス600に属するパーティグループ番号501を示す。同様に、各クラス600の属性は、クラス属性情報  
402に設定されている。本実施の形態におけるクラス  
属性情報402の一例を図13に示す。

【0054】クラス番号は、各クラス600を示す番号である。使用上限値は後述のディスク使用率の許容範囲を示す上限値であり、クラス600の属するパーティ  
50 グループ501に適用する。クラス間性能順位は、クラス600間の性能順位(数字の小さいものが高性能とす  
る)である。クラス間性能順位は各クラス600を構成  
するパーティグループ501の前述の性能差に基づ  
55 く、所記実行上限値および固定値については後述する。  
【0055】本実施の形態における使用状況情報取得処理311について図14で説明する。

【0056】制御部300は、第1の実施の形態と同様に、リード/ライト処理310において使用したディ  
60 スク装置502の使用時間と取得して単位当たりの使

用時間(使用率)を求め、さらに、ディスク装置502  
が属するパーティグループ501について、使用率の平  
均を算出し(ステップ1300)、使用率平均を、リー  
ド/ライト対象となった論理領域についてのディスク装  
55 置使用率として論理領域使用状況情報403に記録する(ステップ1310)。また制御部300は、パーティ  
グループ501に対応する全論理領域のディスク装置使  
用率の和を求め(ステップ1320)、パーティグルー  
プ501の使用率として物理領域使用状況情報404に  
10 記録する(ステップ1330)。

【0057】本実施の形態における論理領域使用状況情報403および物理領域使用状況情報404の一例を図  
15および図16に示す。

【0058】日時とはサンプリング間隔(一定期間)毎の  
日時を示し、論理アドレスは論理領域を示し、パーティ  
15 グループ番号は個々のパーティグループを示し、論理  
領域のディスク装置使用率およびパーティグループ使用率  
はそれぞれ前記サンプリング間隔での平均使用率を示  
す。上記のようなディスク装置502の使用率はディ  
20 スク装置502にかかると示す値であり、使用率が大きい場合は、ディスク装置502が性能ボトルネックとな  
っている可能性があるため、再配置処理で使用率を下  
げることによりディスクアレイシステム201の性能向  
上が期待できる。

【0059】次に、再配置判断処理312について図17で説明する。

【0060】制御部300は、各クラス600について、  
25 クラス600に属するパーティグループ501をク  
ラス構成情報401から取得する(ステップ1300)。  
【0061】続いて、制御部300は、第1の実施の形態と同  
様の再配置判断対象間隔405を参照して対象間隔  
30 を取得し、さらにパーティグループ501について、対  
象間隔の物理領域使用状況情報404のパーティグルー  
プ使用率を取得し集計する(ステップ1320)。扱い  
35 て、制御部300は、クラス属性情報402を参照して  
クラス600の使用率上限値を取得する(ステップ1330)。  
【0062】制御部300は、パーティグループ使用率とク  
ラス上限値と比較し、パーティグループ使用率がク  
ラス上限値より大きい場合は、パーティグループ501の使  
40 用率を減らすために、パーティグループ501に対応す  
る論理領域の再配置が必要と判断する(ステップ1340)。

【0063】扱い、  
45 制御部300は、対象間隔の論理  
領域使用状況情報403を参照して、再配置が必要と判  
断したパーティグループ501の各物理領域に対応する  
論理領域のディスク装置使用率を取得し集計して(ス  
テップ1350)、ディスク装置使用率の大きいものか  
50 60 使用率から選択した論理領域のディスク使用率を減算し

ていき、クラス600の使用率上限値以下になるまで行  
う(1370)。ディスク装置使用率の大きい論理領域  
は、パーティグループ501の使用率に対する影響も大  
きく、またホスト100からの論理領域に対するアクセ  
55 頻度も大きいと考えられるため、ディスク装置使用率  
の大きい論理領域を優先的に再配置することで、ディ  
スクアレイシステム201の効果的な性能改善が期待でき  
る。

【0062】制御部300は、選択された論理領域につ  
いての再配置先となる物理領域を探す。制御部300  
10 は、クラス属性情報402を参照し、パーティグルー  
プ501の属するクラス600より性能順位が高位のクラ  
ス600(高性能クラス)に注目し、クラス構成情報4  
01および第1の実施の形態と同様の未使用領域情報4  
15 0を参照して高性能クラスに属するパーティグループ  
501の未使用物理領域を取得する(ステップ1380)。

【0063】さらに、制御部300は、各未使用物理領  
域について、再配置先とした場合のパーティグループ使  
用率の予測値を求め(ステップ1390)、未使用物理  
領域の中から、再配置先とした場合に高性能クラスに設  
20 定されている上限値を超えないと予測できる未使用物理  
領域を、再配置先の物理領域として選択し(ステップ1  
400)、選択結果を第1の実施の形態と同様に、再配  
置情報408に出力する(ステップ1410)。選択し  
た全ての論理領域について再配置先の物理領域を選択し  
てから処理を終了する(ステップ1420)。

【0064】本実施の形態において、制御部300は、  
第1の実施の形態に加えてパーティグループ情報409  
を保持し、パーティグループ情報409、論理領域使用  
状況情報403、及び物理領域使用状況情報404から  
使用率予測値を算出する。

【0065】パーティグループ情報409の一例を図18に示す。パーティグループ番号は個々のパーティグル  
ープ501を示す番号である。RAID構成はパーティ  
グループ501が構成するRAIDのレベルやディスク  
台数や冗長度構成を示す。ディスク装置性能はパーティ  
グループ501を構成するディスク装置502の性能特  
性(例えば、転送率、アクセス時間、遅延時間)を指し  
40 示す。固定値については後述する。上記の処理にお  
いてディスク装置使用率の大きい論理領域の再配置先を高  
性能クラスのパーティグループ501とすることで、同  
一荷役に対するディスク装置使用時間を短縮でき、論理  
領域の再配置後のディスク装置使用率を抑制できる。

【0066】再配置実行処理313は、第1の実施の形  
態と同様に行われるが、図19に示すように、制御部3  
00は、再配置のためのコピーを行う前にクラス属性信  
50 息402を参照し、再配置先および再配置先のクラス6  
00について、ユーザによってまたは初期条件として設  
定された再配置実行上限値を取得する(ステップ1500)。  
さらに物理領域使用状況情報404を参照して、





なるディスク装置502に分離することができ、特にランダムアクセスに対する応答性を改善することができ、

【0095】また、上記の処理においては制御部300は、シーケンシャルアクセスに注目して再配置による自動的分離を行うことが、同様にランダムアクセスに注目して自動的分離を行うことも可能である。

【0096】上記の再配置処理312において、再配置する論理領域を選択した時点で、制御部300が論理領域属性情報411を参照し、論理領域に固有属性が指定されている場合は、論理領域を再配置しないとすれば、ユーザが特に再配置を行いたくないと考える論理領域がある場合、固定属性を設定することで論理領域を再配置の対象外とすることができ、上記の固定属性に関する処理は論理領域属性情報411を用いることで、前述の実施の形態にも適用できる。

【0097】

【発明の効果】ストレージサブシステムのユーザ、または管理者が、記憶領域の物理的再配置による配置最適化を行うための作業を簡便にすることができる。

【図面の簡単な説明】

【41】本発明の第1の実施の形態での計算機システムの構成図である。

【42】本発明の第1の実施の形態でのリード/ライト処理310および使用状況取得処理311のフローチャートである。

【43】本発明の第1の実施の形態での論理/物理対応情報400の一例を示す図である。

【44】本発明の第1の実施の形態での再配置判断処理312のフローチャートである。

【45】本発明の第1の実施の形態での再配置判断対象期間情報405の一例を示す図である。

【46】本発明の第1の実施の形態での再配置情報408の一例を示す図である。

【47】本発明の第1の実施の形態での未使用領域情報407の一例を示す図である。

【48】本発明の第1の実施の形態での再配置実行処理313のフローチャートである。

【49】本発明の第1の実施の形態での再配置実行時刻情報406の一例を示す図である。

【410】本発明の第2の実施の形態および第5の実施の形態の計算機システムの構成図である。

【411】本発明の第2の実施の形態での論理/物理対応情報400の一例を示す図である。

【412】本発明の第2の実施の形態でのクラス構成情報401の一例を示す図である。

【413】本発明の第2の実施の形態でのクラス属性情報402の一例を示す図である。

【414】本発明の第2の実施の形態での使用状況取得処理311のフローチャートである。

【図15】本発明の第2の実施の形態での論理領域使用状況情報403の一例を示す図である。

【図16】本発明の第2の実施の形態での物理領域使用状況情報404の一例を示す図である。

【図17】本発明の第2の実施の形態での再配置判断処理312のフローチャートである。

【図18】本発明の第2の実施の形態でのパリティグループ情報409の一例を示す図である。

【図19】本発明の第2の実施の形態での再配置実行処理313のフローチャートである。

【図20】本発明の第3の実施の形態での再配置判断処理312のフローチャートである。

【図21】本発明の第4の実施の形態での再配置判断処理312のフローチャートである。

【図22】本発明の第5の実施の形態でのクラス属性情報402の一例を示す図である。

【図23】本発明の第5の実施の形態での論理領域使用状況情報403の一例を示す図である。

【図24】本発明の第5の実施の形態でのアクセス種別基礎値情報410の一例を示す図である。

【図25】本発明の第5の実施の形態での論理領域属性情報411の一例を示す図である。

【図26】本発明の第5の実施の形態での使用状況取得処理311のフローチャートである。

【図27】本発明の第5の実施の形態での再配置判断処理312のフローチャートである。

【符号の説明】

100 ホスト

200 ストレージサブシステム

300 制御部

310 リード/ライト処理

311 使用状況取得処理

312 再配置判断処理

313 再配置実行処理

400 論理/物理対応情報

401 クラス構成情報

402 クラス属性情報

403 論理領域使用状況情報

404 物理領域使用状況情報

405 再配置判断対象期間情報

406 再配置実行時刻情報

407 未使用領域情報

408 再配置情報

409 パリティグループ情報

410 アクセス種別基礎値情報

411 論理領域属性情報

500 記憶装置

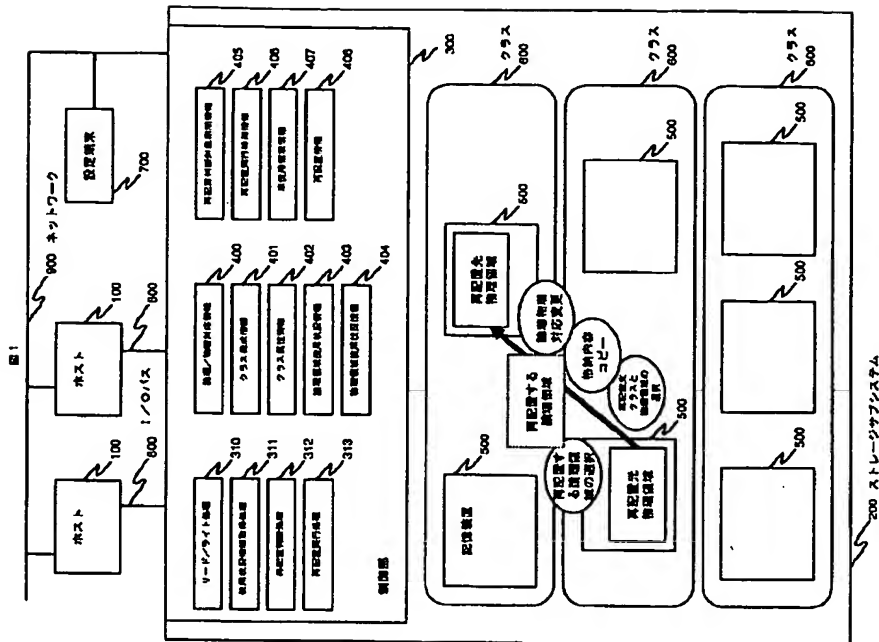
501 パリティグループ

502 ディスク装置

600 クラス  
700 制御部

800 I/Oバス  
900 ネットワーク

【図1】



【図9】

【図24】

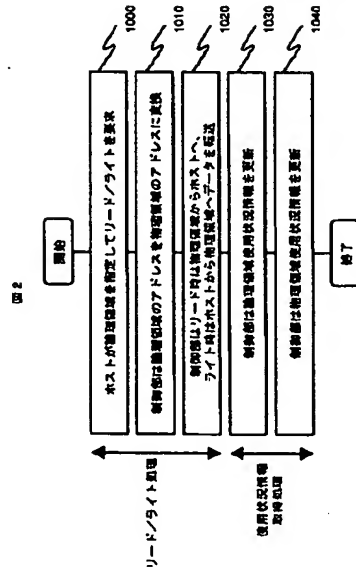
図9

図24

|        |                  |
|--------|------------------|
| 日時     | 1999年8月11日 22時0分 |
| 日時更新情報 | 曜日 (+2.4時間)      |

|              |     |
|--------------|-----|
| アクセス頻度情報 (%) | 7.6 |
|--------------|-----|

【図2】



【図3】

図3

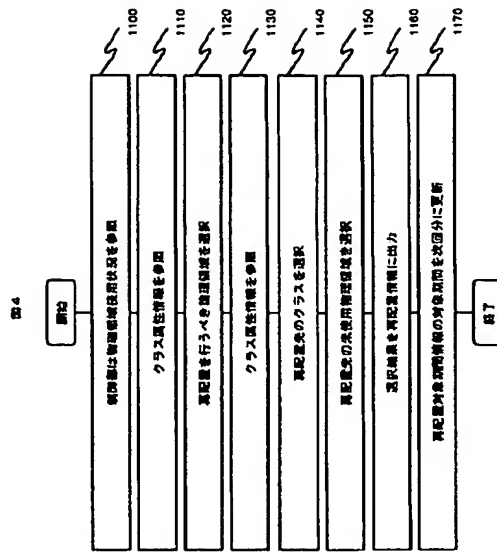
| 記憶アドレス    | 記憶アドレス |           |
|-----------|--------|-----------|
|           | 記憶領域番号 | 記憶領域内アドレス |
| 0~999     | 0      | 0~999     |
| 1000~1999 | 0      | 1000~1999 |
| 2000~2999 | 1      | 0~999     |
| 3000~3999 | 1      | 1000~1999 |

【図5】

図5

|        |                   |
|--------|-------------------|
| 開始日時   | 1999年8月11日 8時30分  |
| 終了日時   | 1999年8月11日 17時15分 |
| 期間更新情報 | 無し (+24時間)        |

【図4】



【図6】

図6

| 番号 | 記憶領域      | 対応属性情報使用状況 |           | 対応属性情報使用状況 |           |
|----|-----------|------------|-----------|------------|-----------|
|    |           | 記憶領域番号     | 記憶領域内アドレス | 記憶領域番号     | 記憶領域内アドレス |
| 1  | 0~999     | 0          | 0~999     | 10         | 0~999     |
| 2  | 1000~1999 | 0          | 1000~1999 | 10         | 1000~1999 |

【図7】

図7

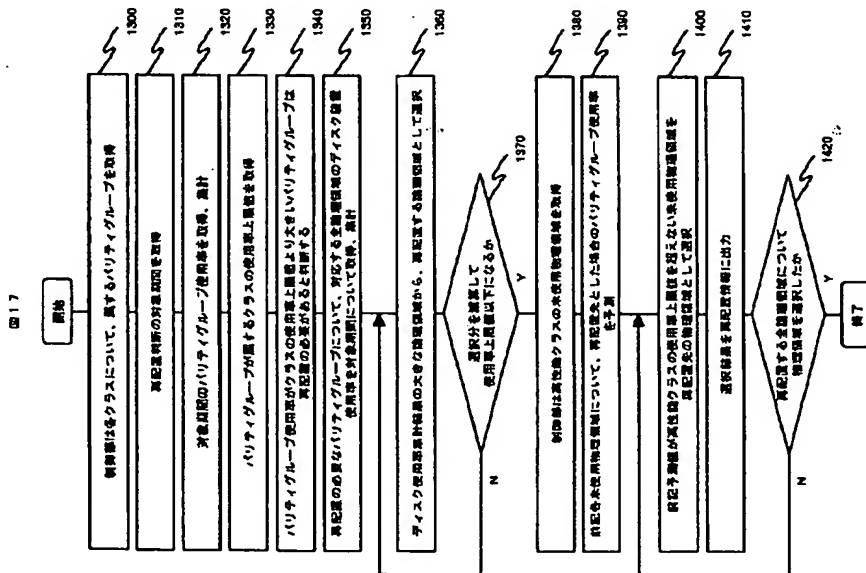
| 記憶領域番号 | 記憶領域内アドレス | 使用/未使用 |
|--------|-----------|--------|
| 0      | 0~999     | 使用     |
| 0      | 1000~1999 | 使用     |
| 0      | 2000~2999 | 未使用    |
| 0      | 3000~3999 | 未使用    |



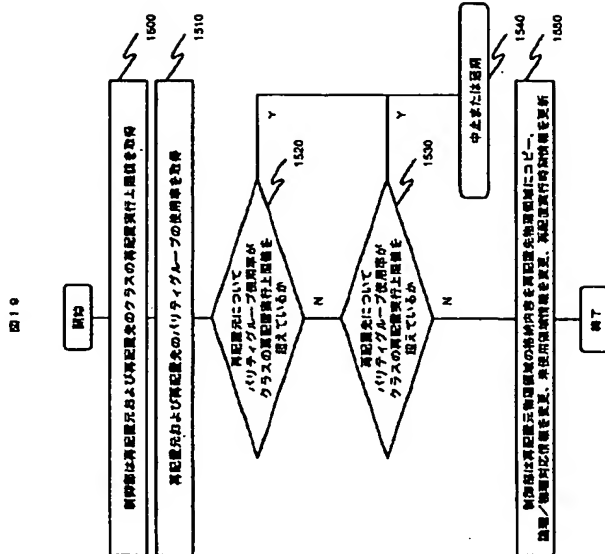




【図17】



【図19】

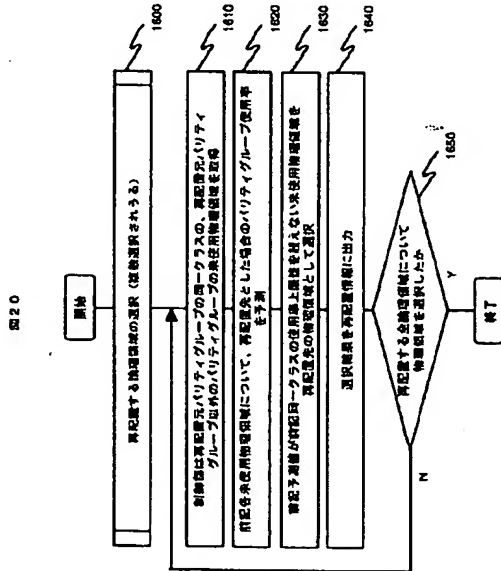


【図23】

図23

| 日時               | 電源アドレス    | ディスク容量<br>使用率 (%) | シーケンシャル<br>アクセス率<br>(%) | ランダム<br>アクセス率<br>(%) |
|------------------|-----------|-------------------|-------------------------|----------------------|
| 1999年8月11日 8時0分  | 0~999     | 18                | 78                      | 22                   |
|                  | 1000~1999 | 32                | 82                      | 40                   |
| 1999年8月11日 8時15分 | 0~999     | 20                | 80                      | 20                   |
|                  | 1000~1999 | 30                | 80                      | 30                   |
| 1999年8月11日 8時30分 | 0~999     | 22                | 82                      | 18                   |
|                  | 1000~1999 | 28                | 40                      | 52                   |

【図20】

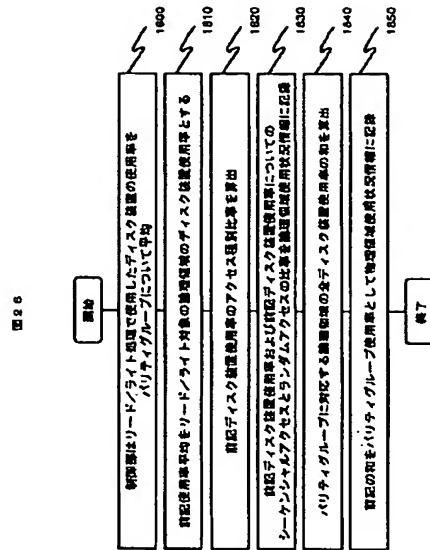


【図25】

図25

| 物理アドレス    | アクセス頻度ヒント | 優先度 |
|-----------|-----------|-----|
| 0~999     | -         | -   |
| 1000~1999 | -         | -   |
| 2000~2999 | シーケンシャル   | -   |
| 3000~3999 | -         | 優先度 |

【図26】



【図21】

